

Das GESIS Datenarchiv für Sozialwissenschaften

Mauer, Reiner

Veröffentlichungsversion / Published Version

Sammelwerksbeitrag / collection article

Zur Verfügung gestellt in Kooperation mit / provided in cooperation with:

GESIS - Leibniz-Institut für Sozialwissenschaften

Empfohlene Zitierung / Suggested Citation:

Mauer, R. (2012). Das GESIS Datenarchiv für Sozialwissenschaften. In R. Altenhöner, & C. Oellers (Hrsg.), *Langzeitarchivierung von Forschungsdaten: Standards und disziplinspezifische Lösungen* (S. 197-215). Berlin: Scivero. <https://nbn-resolving.org/urn:nbn:de:0168-ssoar-46476-7>

Nutzungsbedingungen:

Dieser Text wird unter einer Deposit-Lizenz (Keine Weiterverbreitung - keine Bearbeitung) zur Verfügung gestellt. Gewährt wird ein nicht exklusives, nicht übertragbares, persönliches und beschränktes Recht auf Nutzung dieses Dokuments. Dieses Dokument ist ausschließlich für den persönlichen, nicht-kommerziellen Gebrauch bestimmt. Auf sämtlichen Kopien dieses Dokuments müssen alle Urheberrechtshinweise und sonstigen Hinweise auf gesetzlichen Schutz beibehalten werden. Sie dürfen dieses Dokument nicht in irgendeiner Weise abändern, noch dürfen Sie dieses Dokument für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, aufführen, vertreiben oder anderweitig nutzen.

Mit der Verwendung dieses Dokuments erkennen Sie die Nutzungsbedingungen an.

Terms of use:

This document is made available under Deposit Licence (No Redistribution - no modifications). We grant a non-exclusive, non-transferable, individual and limited right to using this document. This document is solely intended for your personal, non-commercial use. All of the copies of this documents must retain all copyright information and other information regarding legal protection. You are not allowed to alter this document in any way, to copy it for public or commercial purposes, to exhibit the document in public, to perform, distribute or otherwise use the document in public.

By using this particular document, you accept the above-stated conditions of use.

Das GESIS Datenarchiv für Sozialwissenschaften

Reiner Mauer

Das GESIS Datenarchiv für Sozialwissenschaften (DAS)

„Der eigentliche Sinn jedweder langfristigen Archivierung liegt darin, dass einmal erfasste Forschungsdaten erneut genutzt werden können. Die Effekte solcher Nachnutzung werden umso größer sein, wenn Forschungsdaten in dezidierte, in einer Community weithin bekannte und akzeptierte Repositorien eingepflegt und von dort abgerufen werden können“. (Kleiner 2012: 10)

Das Datenarchiv für Sozialwissenschaften stellt Forschungsdaten, vorwiegend aus nationalen und international-vergleichenden Umfragen für die Sekundärnutzung bereit. Die Studien werden gemäß klar definierten methodisch-technischen Anforderungen akquiriert und sodann bedarfsorientiert aufbereitet, dokumentiert, langfristig gesichert und der wissenschaftlich interessierten Öffentlichkeit zugänglich gemacht.

Die Sozialwissenschaften verfügen über eine vergleichsweise lange Tradition in der Archivierung und Nachnutzung von Forschungsdaten. So werden zur Beschreibung und Analyse sozialer Sachverhalte oder zur Entwicklung und Überprüfung sozialwissenschaftlicher Theorien häufig Forschungsdaten herangezogen, die nicht speziell zur Beantwortung der jeweiligen Forschungsfrage erhoben wurden. Oder es werden Daten verwendet, die von vornherein beispielsweise als Mehrthemenbefragung darauf ausgerichtet sind, Analysepotentiale für unterschiedliche Forschungsfragen bereitzustellen. Diese in den Sozialwissenschaften als Sekundäranalyse bezeichnete Nachnutzung von Daten durch nicht unmittelbar an der Datenerhebung beteiligte Dritte (oder auch durch den Primärforscher zur Analyse neuer Forschungsfragen) wurde nicht zuletzt erst durch den Aufbau entsprechender Infrastrukturen ermöglicht und befruchtete diese wiederum. Dies geschah sowohl durch das Angebot auf ein reiches Datenangebot zugreifen zu können als auch durch die gezielte Förderung des data sharing und der Sekundäranalyse als wertvoller Forschungsmethode:

„ ... high priority was assigned to convincing the community of empirical social scientists that a data archive was not a repository for ‘used data’ but a research facility. This meant the promotion of secondary

analysis as an approach that could lead to findings as interesting as those of primary research.” (Scheuch 1990: 96).

Der Aufbau entsprechender Datenserviceinfrastrukturen begann bereits kurz nach dem Zweiten Weltkrieg mit der Gründung des Roper Centers, welches dann allerdings auch erst im Jahr 1957 für eine weitere Öffentlichkeit zugänglich wurde (Scheuch 2003: 386) und entwickelte eine gewisse Dynamik in den 1960er Jahren. Beginnend mit der Etablierung des Zentralarchivs in Köln, folgte eine Reihe von weiteren Gründungen von sozialwissenschaftlichen Datenarchiven in den Folgejahren (wie beispielsweise dem ICPSR in Ann Arbor, dem Steinmetz Archiv in den Niederlanden und dem UK Data Archive in Colchester). Ein weiterer Schub an Gründungen war dann in den 1990er Jahren zu verzeichnen, insbesondere in den Staaten Mittel- und Osteuropas. Allein in Europa sind gegenwärtig 21 sozialwissenschaftliche Datenarchive im Council of European Social Science Data Archives (CESSDA¹) organisiert. Dieses 1976 gegründete Kooperationsnetzwerk nationaler Archive entwickelt sich gegenwärtig im Rahmen des ESFRI-Prozesses zu einem sogenannten European Research Infrastructure Consortium (ERIC) und damit zu einer Körperschaft europäischen Rechts, die dann auch dauerhaft koordinierende Funktionen für die Datenlandschaft in Europa übernehmen kann (Quandt und Mauer 2012: 66).

Die in der jüngeren Vergangenheit enorm gewachsene Aufmerksamkeit hinsichtlich der Bedeutung von Forschungsdaten hat auch in den Sozialwissenschaften zu einer erneuten Dynamik beim Aufbau von Forschungsdateninfrastrukturen geführt. In Deutschland sind dabei insbesondere die Aktivitäten des Rates für Sozial- und Wirtschaftsdaten (RatSWD²) und die in seinem Umfeld entstandenen Forschungsdatenzentren (FDZ) zu nennen.

Institutionelle Verankerung

1960 als Zentralarchiv für Empirische Sozialforschung an der Universität zu Köln gegründet, ist das Archiv heute eine von fünf wissenschaftlichen Abteilungen von GESIS – Leibniz-Institut für Sozialwissenschaften, die mit ihrem forschungsbasierten Service- und Produktangebot den Forschungsprozess der empirischen Sozialforschung in seiner gesamten Breite abdecken. Dabei dient der Forschungsdatenzyklus als Leitbild zur Strukturierung und Verknüpfung der Angebote. GESIS ist Mitglied der Leibniz-Gemeinschaft und erbringt mit seinen über 250 Mitarbeitern an drei Standorten (Mannheim, Köln, Berlin) grundlegende, überregional und international bedeutsame forschungsbasierte Dienstleistungen.

¹ <http://www.cessda.org>

² <http://www.ratswd.de>

Leitgedanken bei der Gründung

Die Untersuchung des sozialen Wandels, also beispielsweise von politischem, technologischem oder ökonomischem Wandel oder von auch Veränderungen im Wertesystem, ist eines der klassischen Themen sozialwissenschaftlicher Forschung. Sollen derartige Veränderungen im Zeitverlauf empirisch untersucht werden, besteht die Notwendigkeit auf bereits erhobene Daten zurückzugreifen, da es nicht immer oder nur bedingt möglich ist, entsprechende Daten retrospektiv zu erheben, die beispielsweise Auskunft über Einstellungen oder Verhalten in der Vergangenheit geben können. Die Archivierung sozialwissenschaftlicher Daten und die damit häufig erst mögliche Nachnutzung, „... extend[s] any particular survey from a mere snapshot of reality into a continuous observation.“ (Scheuch 1990: 96).

Neben diesem zentralen methodischen Motiv spielten forschungsökonomische Argumente eine gleichsam bedeutende Rolle bei der Gründung. Insbesondere die Beobachtung, dass mit vergleichsweise hohem finanziellem und zeitlichem Aufwand erhobene Daten in der Regel nur sehr begrenzt in einem einzigen Forschungskontext analysiert wurden, häufig aber auch über darüber hinausgehende, bisher nicht ausgeschöpfte Analysepotentiale verfügen. Gleichfalls sollte die Möglichkeit, auf bereits erhobene Daten zurückzugreifen, zur Vermeidung unnötiger Doppelungen beitragen und Forschende in stärkerem Ausmaß dazu anregen, neue Erhebungen auf der Basis bestehender Arbeiten aufzubauen, um somit auch eine größere Vergleichbarkeit von Forschungsergebnissen zu erreichen. Ganz allgemein zielte die Gründung des Datenarchivs darauf ab, den Zugang zu Forschungsdaten – gerade auch für Studierende und Nachwuchswissenschaftler – zu erleichtern und auf diesem Wege die empirische Sozialforschung und im Speziellen Sekundäranalysen sowie Methodenforschungen zu unterstützen.

Aufgaben

Zu den zentralen Aufgaben des GESIS Datenarchivs zählt insbesondere die Langzeitarchivierung von Forschungsdaten, also die langfristige Sicherstellung von Interpretierbarkeit und Verfügbarkeit der archivierten Bestände und die Bereitstellung dieser Daten. Dies geschieht mit dem Ziel, Sekundäranalysen ganz allgemein und im Besonderen Raum- und Zeitvergleiche zu ermöglichen. Des Weiteren leistet die Arbeit des Archivs einen wichtigen Beitrag zur Erhöhung von Transparenz und Überprüfbarkeit wissenschaftlicher Forschung, indem es Replikationen und Re-Analysen ermöglicht.³

³ Wenngleich, wie Wagner und Huschka (2012) sicher zu Recht anmerken, die Überprüfung von Forschungsergebnissen auch in den Sozial- und Wirtschaftswissenschaften nicht gerade eine vorherrschende Praxis ist und die bloße Verfügbarkeit von Forschungsdaten allein nicht ausreicht, um Replikationen zu fördern. Jedoch gilt andersherum natürlich auch, dass ohne die Zugriffsmöglichkeit auf die fraglichen Daten Replikationen erst gar nicht möglich sind.

Eine weitere wichtige Aufgabe des GESIS Datenarchivs besteht darin, den Zugang zu internationalen Forschungsdaten zu erleichtern. Diese geschieht einerseits durch die Archivierung ausländischer Datenbestände⁴ und hierbei insbesondere durch die Betreuung großer interkulturell und intertemporal vergleichender Umfrageprogramme, wie etwa die Eurobarometer Studien der Europäischen Kommission oder die Daten der Europäischen Wertestudie (EVS)⁵. Zum anderen aber auch durch die Vermittlung von Daten, die bei anderen Archiven gehalten werden. Als Mitglied in internationalen Archivnetzwerken wie IFDO (International Federation of Data Organizations for Social Science⁶) und dem europäischen Archivverbund CESSDA (Council of European Social Science Data Archives⁷) sowie als nationaler Repräsentant des ICPSR (Inter-university Consortium for Political and Social Research⁸) vermittelt das GESIS Datenarchiv für deutsche Forscher den Zugang zu sozialwissenschaftlichen Forschungsdaten weltweit.

Ferner unterstützt das Datenarchiv Primärforscher bei der Sicherung, Dokumentation, Aufwertung und Bereitstellung ihrer Daten. Sei es durch die Beratung in allen Fragen des Datenmanagements, aber auch durch praktische Unterstützung der Arbeit oder die Bereitstellung von Tools.

Datenbestand

Der Datenbestand umfasst gegenwärtig knapp 5.100 öffentlich zugängliche Studien. Dabei handelt es sich zum größten Teil um Mikrodaten der Umfrageforschung sowie um historische Zeitreihen und Aggregatdaten. Thematisch liegt der Fokus auf Studien zu sozial- und politikwissenschaftlichen Fragestellungen. Ein wichtiger und intensiv betreuter Schwerpunkt sind dabei Daten der interkulturell vergleichenden Sozialforschung, die eine Vielzahl von Ländern und überwiegend lange Zeiträume abdecken – wie etwa das International Social Survey Programme (ISSP), das jährlich eine gemeinsame Umfrage zu sozialwissenschaftlich relevanten Themen durchführt und seit seiner Gründung 1984 auf 48 Mitgliedsländer im Jahre 2011 angewachsen ist⁹. Auch kontinuierliche nationale Erhebungsprogramme, wie beispielsweise die seit 1980 von GESIS bzw.

4 So hält das Datenarchiv beispielsweise eine größere Sammlung von Daten aus Mittel- und Osteuropa oder die Daten des Japanese General Social Survey.

5 <http://www.europeanvaluesstudy.eu>

6 <http://www.ifdo.org>

7 <http://www.CESSDA.org>

8 www.icpsr.umich.edu

9 <http://www.issp.org> bzw. <http://www.gesis.org/issp/>. Weitere wichtige von GESIS (teilweise in Kooperationen) betreute internationale Umfrageprogramme sind etwa die Eurobarometer Studien der Europäischen Kommission, die European Values Study (EVS) oder auch die Comparative Study of Electoral Systems (CSES), die in rund 40 Ländern politische Einstellungen und Wahlverhalten untersucht. Eine Übersicht und weiterführende Hinweise finden sich unter <http://www.gesis.org/das-institut/kompetenz-zentren/fdz-internationale-umfrageprogramme/>

seinen Vorgängerinstituten durchgeführte Allgemeine Bevölkerungsumfrage der Sozialwissenschaften (ALLBUS)¹⁰ oder die erstmals anlässlich der Bundestagswahl 2009 durchgeführte German Longitudinal Election Study (GLES)¹¹ werden aufwendig betreut. Darüber hinaus hält das GESIS Datenarchiv umfangreiche Bestände zu politischen Einstellungen und Verhalten, Werten, Jugend, Gesundheit, Mediennutzung und vielen weiteren sozialwissenschaftlich relevanten Themenbereichen vor.

Datenarchivierung@DAS: Konzeptionell

Auf der Grundlage des 1999 veröffentlichten OAIS Referenzmodells entwickelte sich ein zunehmendes Bewusstsein und mittlerweile weithin geteiltes Verständnis für die Notwendigkeit von archivarischen Konzepten, die darauf abzielen, den langfristigen Erhalt von und den Zugang zu digitalen Informationen sicherzustellen. Mit dem mittlerweile als ISO-Standard etablierten Referenzmodell wurde sowohl eine Nomenklatur als auch ein Rahmen bereitgestellt, der es ermöglicht, unterschiedlichste (digitale) Archive und deren Archivierungskonzepte in einer einheitlichen Form zu beschreiben und darauf aufbauend ihre Funktionsfähigkeit und auch Vertrauenswürdigkeit beurteilen zu können.

Entsprechend dem damaligen technologischen Stand waren die ersten Jahre der Datenarchivierung in Köln verständlicherweise nicht ‚digital‘ geprägt, sondern wurden von der Notwendigkeit bestimmt, auf Lochkarten vorgehaltene Forschungsdaten zu verarbeiten. Die Sicherung der Forschungsdaten erfolgte durch Herstellung und Aufbewahrung von Duplikaten dieser Lochkarten – also auf einem Speichermedium aus Papier – und die Datenverarbeitung beschränkte sich weitgehend auf das Auszählen und Sortieren der Karten mittels sogenannter Fachzählsortiermaschinen. Dies änderte sich jedoch relativ bald durch den Einsatz von Großrechnern (zunächst extern und dann mit eigenem Rechenzentrum) und entwickelte große Dynamik mit dem Aufkommen von Mikrocomputern, so dass man spätestens seit den siebziger Jahren des letzten Jahrhunderts das Datenarchiv auch als digitales Langzeitarchiv bezeichnen kann.

Der sich am OAIS orientierende Kriterienkatalog vertrauenswürdige digitale Langzeitarchive von nestor (2006: 2) definiert ein digitales Langzeitarchiv als eine *„Organisation [...], die die Verantwortung für den Langzeiterhalt und die Langzeitverfügbarkeit digitaler Objekte sowie für ihre Interpretierbarkeit*

10 <http://www.gesis.org/allbus>

11 <http://www.gles.eu/> bzw. <http://www.gesis.org/wahlen/gles/>. Darüber hinaus betreut GESIS auch die vom ZDF beauftragte und der Forschungsgruppe Wahlen seit 1977 monatlich durchgeführten Politbarometer sowie die im Auftrag der ARD erhobenen DeutschlandTrends. Siehe hierzu: <http://www.gesis.org/wahlen/>

zum Zwecke der Nutzung durch eine bestimmte Zielgruppe übernommen hat.“ Die beiden wesentlichen Bestandteile dieser Definition, nämlich die Übernahme der Verantwortung für die Langzeitarchivierung von digitalen Objekten und die Verfügbarkeit dieser Objekte – im Fall des Datenarchivs von Forschungsdaten – sowie die Zielgruppenorientierung (Sozialwissenschaften) sind explizit in der Satzung von GESIS¹² verankert. So bestimmt § 2, mit welchen konkreten Aufgaben der Vereinszweck – nämlich die „Förderung der sozialwissenschaftlichen Forschung“¹³ – erfüllt wird. Dazu zählen unter anderem die „c) Archivierung, Dokumentation und Langzeitsicherung sozialwissenschaftlicher Daten, einschließlich ihrer Erschließung sowie qualitativ hochwertigen Aufbereitung besonders relevanter Daten für Sekundäranalysen, ...“¹⁴ sowie die „e) Schaffung eines benutzerfreundlichen und hochqualitativen Zugangs zu allen für die empirische Sozialforschung relevanten Informationen und Daten ...“¹⁵.

Um eine möglichst breite und optimale Nachnutzung zu ermöglichen, steht das aktive Management der Daten und insbesondere die Schaffung von Mehrwert durch Aufbereitung, Dokumentation und die Verknüpfung von Daten seit den Gründungstagen im Mittelpunkt der Arbeit des Datenarchivs. Dieses Verständnis von Datenarchivierung bei GESIS lässt sich in aktueller Terminologie am besten mit dem Begriff ‚data curation‘ beschreiben:

“Digital curation is all about maintaining and adding value to a trusted body of digital information for future and current use; specifically, the active management and appraisal of data over the entire life cycle. [...] builds upon the underlying concepts of digital preservation whilst emphasising opportunities for added value and knowledge through annotation and continuing resource management [...].” (JISC 2006: 1)

Diese Schwerpunktsetzung erfolgte bereits bei Gründung in bewusster Abgrenzung zu den Erfahrungen, die Scheuch – einer der beiden Gründungsväter des Archivs – beim Besuch des Roper Centers in den fünfziger Jahren machte. Heute eines der bedeutendsten sozialwissenschaftlichen Archive, war es damals nicht zuletzt aufgrund der fehlenden finanziellen Ausstattung mehr oder weniger eine reine Lagerstätte für Lochkarten (Scheuch 1990: 95).

12 GESIS – Leibniz-Institut für Sozialwissenschaften e.V., Satzung vom 19.06.2010. <http://www.gesis.org/das-institut/der-verein/satzung/> [21.05.2012]

13 Ebd., § 2 (1), Satz 1

14 Ebd., § 2 (2) (c)

15 Ebd., § 2 (2) (e)

Datenarchivierung@DAS: Funktional

Die zuvor erfolgte konzeptionelle Einordnung des Datenarchivs soll im Folgenden um eine Beschreibung der konkreten Arbeitsweisen des Archivs erweitert werden (die aus Platzgründen naturgemäß nur einen groben Überblick über die Abläufe geben kann). Am Anfang der Archivierungstätigkeit steht die Akquisition von Studien, also die Entscheidung darüber, welche Daten überhaupt in das Archiv aufgenommen werden sollen. Daran schließt sich die konkrete Übernahme der Forschungsdaten in das Archiv an (Ingest). In der Folge durchlaufen die Studien verschiedene Aufbereitungs- und Dokumentationsschritte, wobei es einen gemeinsamen Standard gibt, den alle Studien durchlaufen, und darüber hinausgehende aufwendigere Prozesse, die nur für ausgewählte bzw. besonders betreute Studien durchgeführt werden. Im Anschluss an diese werden Daten und Dokumentation an den Datenservice sowie an die Langzeitarchivierung übergeben.¹⁶

Akquisition

Forschungsdaten – im Regelfall Daten der empirischen Umfrageforschung – gelangen auf unterschiedlichen Wegen ins Archiv: Sie werden entweder gezielt akquiriert oder aber dem Archiv von Primärforschern zur Archivierung angeboten bzw. erfolgt die Übergabe regelmäßig im Rahmen längerfristiger Kooperationen. Grundsätzlich werden maschinenlesbare Datensätze aus allen Bereichen der Sozialwissenschaften archiviert, sofern (1) die Studien Aussagen über die deutsche Bevölkerung oder über Teile von ihr erlauben, (2) an der Untersuchung deutsche Forscher beteiligt waren, unabhängig davon, ob sie sich auf Deutschland bezieht oder nicht, und (3) die Studie ganz allgemein für die sozialwissenschaftliche Forschung von Interesse sein könnte. Gezielt akquiriert werden dagegen meist Daten zu bestimmten Forschungsgebieten (beispielsweise der international vergleichenden Sozialforschung), besonders prominente und bedeutende Erhebungsprojekte oder solche, die bereits im Archiv vorhandene Kollektionen vervollständigen. Neben den zuvor genannten Kriterien, müssen die aufzunehmenden Studien bestimmten formalen und technischen Anforderungen entsprechen. Insbesondere müssen neben den Datensätzen selbst, auch alle für eine Sekundärnutzung notwendigen Materialien vorhanden sein und an das Archiv übergeben werden (Erhebungs- bzw. Messinstrumente, Codepläne, Methodenberichte etc.). Diese den Datensatz begleitenden Materialien und Dokumente bilden auch die Grundlage für die vom Datenarchiv im Rahmen der Archivierung erzeugten standardisierten Metadaten.

¹⁶ Für den ebenfalls von GESIS im Rahmen von da|ra angebotenen Datenregistrierungsservice sei auf den Beitrag von Brigitte Hausstein in diesem Band verwiesen.

Um eine Archivierung und damit insbesondere eine Nachnutzung durch Dritte optimal vorzubereiten, ist es hilfreich, diese möglichst frühzeitig einzuplanen. Idealerweise beginnt dies bereits bei der Vorbereitung einer Datenerhebung. Insofern ist es auch das Bestreben des Datenarchivs, möglichst früh im Lebenszyklus einer Studie anzusetzen. Auch wenn sich langsam das Wissen um die Vorteile einer proaktiven Planung verbreitet – nicht zuletzt durch die Diskussion um die Notwendigkeit von Datenmanagementplänen –, so stellt sie doch derzeit (noch) nicht den Regelfall in der Archivierungspraxis dar. Vielmehr werden Studien häufig erst lange nach der Erhebung bzw. nach dem Projektende zur Archivierung angeboten bzw. für diese freigegeben. Dies verursacht je nach Zustand von Daten und Dokumentation einen teils erheblichen Mehraufwand.

Ingest – Aufnahme ins Archiv

Vor der eigentlichen Aufnahme einer Studie ins Archiv, wird diese zunächst gemeinsam mit dem Datengeber vorbereitet. Dazu werden Absprachen über Umfang, Formate, Aufbereitungs- und Dokumentationsziele getroffen sowie geklärt, unter welchen Bedingungen die Daten weitergegeben werden. Diese Absprachen werden schließlich in Form einer Archivierungsvereinbarung festgehalten, in der auch die für eine Archivierung und Weitergabe notwendige Übertragung von Nutzungsrechten an das Archiv festgehalten wird.

Entsprechen die aufzunehmenden Studien den oben genannten Anforderungen und liegt eine entsprechende Archivierungsvereinbarung vor, werden sie beim Dateneingang einer standardisierten Eingangskontrolle unterzogen. Dabei besteht das Submission Information Package bzw. Übergabeinformationspaket (SIP)¹⁷ in der Regel aus einem oder mehreren Datensätzen, den dazugehörigen Messinstrumenten (Fragebogen), sowie weiteren Materialien mit Informationen zum Forschungsprozess und Publikationen. Diese Eingangskontrolle umfasst dabei unter anderem folgende Schritte:

- Technische Kontrollen: Formate, Lesbarkeit, Virenfreiheit etc.
- Vollständigkeit und Nutzbarkeit: Sind Daten, Messinstrumente und Dokumente vollständig übergeben worden und beziehen sie sich aufeinander? Insbesondere: Passen Erhebungsinstrument und Daten zusammen?
- Konsistenz: Gibt es Werte außerhalb des zulässigen Bereichs (wild codes)? Sind fehlende Werte definiert und wenn ja, wie? Stimmen die Daten mit der im Fragenbogen vorgegebenen Filterführung überein (question routing)? Gibt es widersprüchliche Merkmalskombinationen?

¹⁷ „Ein Informationspaket, das vom Produzenten an das OAIS geliefert wird, um es zur Konstruktion oder zur Aktualisierung eines oder mehrerer AIPs und/oder den dazugehörigen Erschließungsinformationen zu benutzen“ (nestor 2012: 15).

- **Datenschutz:** Da es sich bei den archivierten Studien vorrangig um Mikrodaten handelt, werden sie darüber hinaus auch auf datenschutzrechtlich relevante Aspekte untersucht.

Bei der Eingangskontrolle festgestellte Inkonsistenzen oder sonstige Fehler werden dokumentiert und, sofern möglich, nach Rücksprache mit den Datengebern korrigiert.

In Abhängigkeit vom Ergebnis der Eingangskontrolle und einem zuvor festgelegten Aufbereitungs- und Dokumentationsziel folgen der Eingangskontrolle weitere Arbeitsschritte. So werden auf Datenträgern gelieferte Daten und Materialien von den Original-Medien getrennt. Digitale Objekte werden in definierte Langfristsicherungsformate überführt, Materialien im Papierformat werden digitalisiert und die Originale in einem speziell dafür vorgesehenen Archivraum abgelegt. Sodann werden beschreibende Metadaten in Form einer sogenannten Studienbeschreibung angelegt (siehe unten) sowie weitere interne (administrative und technische) Metadaten in einer zentralen Datenbank erzeugt. Ein Teil der eingehenden Studien wird über die Standardarchivierung hinaus einer umfassenderen Datenaufbereitung und -dokumentation zugeführt (siehe unten). Die Ergebnisse dieser Aufbereitungs- und Dokumentationsarbeiten werden nach Abschluss wiederum in die Langfristsicherung übergeben. Schlussendlich werden die für den Datenservice bestimmten Objekte in ein oder mehrere Dissemination Information Packages oder Auslieferungsinformationspakete (DIP)¹⁸ überführt. Dabei erhalten alle Datensätze eine standardisierte Versionsinformation und werden mit einem persistenten Identifikator in Form eines DOI versehen. Die dafür notwendige Registrierung erfolgt bei der von GESIS betriebenen Registrierungsagentur da|ra.

Datenaufbereitung / -dokumentation

Ein wesentlicher Teil der Ressourcen des Archivs wird eingesetzt, um Forschungsdaten so aufzubereiten und zu dokumentieren, dass sie durch Dritte – also nicht unmittelbar an der Datenerhebung beteiligte – genutzt werden können. Der Umfang dieser Arbeiten ist einerseits abhängig vom Ausgangszustand der übernommenen Daten und Materialien und andererseits vom konkreten Aufbereitungs- und Publikationsziel der jeweiligen Studie. Prozesse, die alle Studien durchlaufen, sind aufbauend auf der oben beschriebenen Eingangskontrolle einfache Aufbereitungen (beispielsweise zur Fehlerkorrektur oder zur Herstellung eines vollständig gelabelten Analysedatensatzes), Versionierung und die Zuweisung eines entsprechenden DOI-Namens, sowie die Erstellung einer Studienbe-

¹⁸ „Ein Informationspaket, abgeleitet aus einem oder mehreren AIPs, das der Endnutzer als Antwort auf eine Anfrage an das OAI-S erhält.“ (nestor 2012: 9)

schreibung, die in standardisierter Form (kompatibel zur Metadatenpezifikation der internationalen Data Documentation Initiative, DDI) inhaltliche, methodische und technische Charakteristika einer Studie spezifiziert.¹⁹ Sie ermöglicht externen Nutzern das Auffinden der Studien sowie die Beurteilung der Relevanz für die eigene Forschung. Diese Studienbeschreibungen werden sowohl in deutscher als auch in englischer Sprache in einer Datenbank gepflegt, die online recherchierbar Nutzern zur Verfügung gestellt wird (siehe unten zum Datenbestandskatalog).

Adding value

Um eine möglichst breite und optimale Nachnutzung zu ermöglichen, werden für die Forschung besonders relevante Daten mit teils erheblichem Aufwand mit Mehrwert versehen. Diese Studien durchlaufen zusätzliche teils sehr intensive Aufbereitungs- und Dokumentationsschritte. Dabei handelt es sich hauptsächlich um solche Studien, die sich für die interkulturell vergleichende Forschung eignen und/oder kontinuierlich durchgeführt werden und sich somit grundsätzlich in Zeitreihen einordnen lassen. Für diese meist großen Umfrageprogramme bietet das GESIS Datenarchiv eine auf die jeweiligen Bedarfe maßgeschneiderte Unterstützung im Datenmanagement. Häufig ist GESIS selbst in verschiedenen Funktionen an der Datenerhebung dieser Programme beteiligt²⁰ oder führt diese, wie im Fall der Allgemeinen Bevölkerungsumfrage der Sozialwissenschaften selbst durch. Service und Forschung zu diesen langjährigen GESIS-Aktivitäten wurden mittlerweile in Forschungsdatenzentren gebündelt, die sich an den Kriterien des Rates für Sozial- und Wirtschaftsdaten (RatSWD) orientieren.²¹ Diese bieten Forschern nicht nur hochwertig aufbereitete und dokumentierte Daten an, sondern bedienen auch den besonderen Bedarf an Beratung und Wissensvermittlung, der mit der Nutzung derartiger Daten einhergeht.

Die derart betreuten Daten werden je nach Zielvorgabe weitergehend standardisiert und aufbereitet, harmonisiert und zu komplexen zeit- und/oder ländervergleichenden Studien integriert. Die Dokumentation der Studien wird je nach Zielvorgabe erweitert um Metadaten sowohl auf Studien- als auch auf Datensatzebene (Standardisierung der Datensatzstruktur, Variablenamen/-label, etc.). Darüber hinaus erfolgt eine strukturierte Dokumentation des Fragebogens, ergänzt um Anmerkungen zur Datenqualität auf Variablenebene und weiteren Kontextinformationen. Dieser umfangreiche Bestand an strukturierten Metadaten wird dann einerseits über verschiedene (Meta)Datenportale und Recherche-

19 Zum Metadatenchema der Studienbeschreibungen im Datenbestandskatalog siehe Zenk-Möltgen und Habbel (2012).

20 So ist GESIS beispielsweise am International Social Survey Programme nicht nur für die Datenarchivierung zuständig, sondern – vertreten durch die Abteilung Dauerbeobachtung der Gesellschaft – auch für die Durchführung der deutschen Teilerhebung und übt darüber hinaus weitere internationale Funktionen insbesondere im Methodenbereich aus.

21 <http://www.gesis.org/das-institut/kompetenzzentren/>

systeme, wie etwa dem Datenportal ZACAT, zur Verfügung gestellt und dient andererseits aber auch dazu, begleitende Dokumentationen etwa in der Form von Codebüchern, Variablenreports oder Methodenberichten zu erstellen. Diese sind insbesondere bei komplexeren Studien oder Studienkollektionen neben den Daten selbst und den dazugehörigen Messinstrumenten häufig Voraussetzung für eine sekundäranalytische Nutzung durch Dritte.

Access / Datenbereitstellung

Ein zentraler Bestandteil der Archivierungstätigkeit ist die Bereitstellung der Forschungsdaten für die Nachnutzung, sei es für Replikationszwecke oder aber – und das gilt für die überwältigende Anzahl der Fälle – um neue Forschungsfragen mithilfe dieser Daten zu beantworten. Mit mehr als 30.000 Datenweitergaben in 2011 und der Bearbeitung von mehr als tausend Anfragen rund um das Datenangebot pro Jahr wird der Service des GESIS Datenarchivs intensiv genutzt – weit überwiegend für die akademische Forschung und Lehre. Dabei beschränkt sich die Nutzung nicht auf Deutschland, sondern je nach Angebot werden die Dienstleistungen zwischen 30 % und 70 % von Nutzern außerhalb Deutschlands in Anspruch genommen. Die Bereitstellung der archivierten Daten und Dokumente wird durch vom Datengeber vergebene Zugangskategorien geregelt.²² Rund Dreiviertel der archivierten Studien ist dabei frei zugänglich für die akademische Forschung und Lehre, für ein Viertel der Studien ist vor einer Weitergabe eine schriftliche Genehmigung durch den Datengeber einzuholen. Daten und zugehörige Materialien, wie beispielsweise Fragebögen oder Variablenreports, werden entweder kostenfrei zum Download in verschiedenen Portalen und Systemen angeboten oder können über den Datenservice bestellt werden. Für die individuelle Bereitstellung von Daten wird dann allerdings eine Bereitstellungsgebühr in Rechnung gestellt.

Einen zentralen Zugangspunkt zu den Beständen des Archivs bietet der Datenbestandskatalog, der zu allen archivierten Studien ausführliche Informationen in der Form von Studienbeschreibungen bereitstellt.²³ Neben bibliographischen Angaben, werden insbesondere Inhalte und methodische Aspekte beschrieben, wie etwa die Stichprobenziehung oder die in der Studie eingesetzten Erhebungsverfahren. Darüber hinaus finden sich auch Hinweise auf Veröffentlichungen und die Versionsgeschichte, sowie aktuell bekannte Fehler in den Daten. Neben der Möglichkeit einfache oder auch verknüpfte Suchen durchzuführen, bietet der Datenbestandskatalog die Möglichkeit Daten und Dokumente dieser Studien direkt herunterzuladen oder über ein Warenkorbsystem zu bestellen.

22 http://www.gesis.org/unser-angebot/daten-analysieren/datenservice/benutzungsordnung/#3_Zugangskategorien

23 <http://www.gesis.org/dbk>

Neben dem zentralen Zugang über den Datenbestandskatalog existieren weitere, spezialisierte Angebote, um beispielsweise erweiterte Dokumentation oder zusätzliche Funktionen für die Nutzer anbieten zu können. So bietet das Datenportal ZACAT²⁴ einen direkten Zugang zu gegenwärtig rund 600 Studien. Der Schwerpunkt des Angebotes liegt dabei auf international vergleichenden Studien sowie auf Daten der Wahlforschung, unter anderem Eurobarometer, ISSP, European Values Study, ALLBUS, Politbarometer und Wahlstudien (Deutschland, Osteuropa). Zu jeder Studie wird eine ausführliche Dokumentation sowohl auf Studien- als auch auf Variablenebene, inklusive der vollständigen Frage- und Antworttexte des Fragebogens, angeboten. Für einige Studien sind Frage- und Antworttexte auch mehrsprachig dokumentiert. So sind beispielsweise für die vierte Welle der European Value Study alle in den 49 Teilnahmeländern eingesetzten originalsprachlichen Fragebögen dokumentiert. Neben der Suche in den strukturierten Metadaten auf Studienebene können mithilfe der erweiterten Suche auch die kompletten Frage- und Antworttexte des Fragebogens sowie Variablen- und Wertelabel durchsucht werden und somit aus dem Bestand von weit über 200.000 Variablen einzelne Indikatoren gezielt identifiziert werden. Darüber hinaus ermöglicht das System (einfache) Analysen wie Häufigkeitsauszählungen, Kreuztabellen, Regressionen und deren graphische Darstellung (Balken- und Kuchendiagramme). Die Analysen können auch unter Verwendung von GewichtungsvARIABLEN durchgeführt werden. Die Datensätze selbst sind in verschiedenen gängigen Formaten herunterladbar (unter anderem SPSS, SAS, Stata, csv). Für den Download können auch subsets von Datensätzen gebildet werden (Auswahl von Variablen oder Fällen). Der Zugang ist gebührenfrei, Analyse und Download erfordern eine Registrierung.

Studien mit Zeitreihen bzw. mit zeitreihenfähigen Daten zur historischen Demografie, zur empirischen Sozial- und Wirtschaftsgeschichte sowie zur Historischen Statistik Deutschlands werden mit der Datenbank HISTAT²⁵ bereitgestellt. Die Datenbank ist thematisch untergliedert (zum Beispiel Erwerbstätigkeit, Bevölkerung, Bildung und Wissenschaft etc.) und bietet derzeit Zugang zu rund 260.000 Zeitreihen. Neben diesen Zeitreihen selbst, enthält die Datenbank ausführliche Studienbeschreibungen, die Auskunft über inhaltliche und methodische Aspekte der Studien geben sowie die verwendeten Quellen beschreiben. Neben verschiedenen Recherchemöglichkeiten erlaubt das auf Zeitreihen spezialisierte System auch den Download der Studien bzw. Zeitreihen im Excel- oder CSV-Format.

24 <http://zacat.gesis.org>

25 <http://www.gesis.org/unser-angebot/daten-analysieren/daten-historische-studien/zeitreihen/>

Mit SIMon²⁶, dem Social Indicators Monitor, stellt GESIS ein weiteres Informationssystem mit Zeitreihendaten zur Verfügung. Angeboten werden deutsche und europäische Sozialindikatoren, mit deren Hilfe sozialstruktureller Wandel und die Entwicklung der Lebensbedingungen und Lebensqualität der Bevölkerung beobachtet und analysiert werden kann. SIMon erlaubt neben einer auf verschiedenen Wegen unterstützten Auswahl und Recherche von Indikatoren, einfache Möglichkeiten der Tabellenmanipulation und Datenanalyse sowie die grafische Darstellung der Indikatoren, etwa als Liniengrafik, Histogramm oder Boxplot oder auf Landkarten. Ebenso wird der Export von Tabellen und Grafiken unterstützt, so dass diese in anderen Anwendungen weiterverarbeitet werden können. Über die oben beschriebenen Portale und Systeme hinaus, bietet das Web-Angebot der GESIS eine Fülle an weiteren Informationen rund um das Datenangebot, insbesondere zu den besonders betreuten Umfrageprogrammen bzw. einzelnen Studien in diesen Serien.²⁷

Langzeitarchivierung

Seit seiner Gründung im Jahr 1960 ist die langfristige Sicherung von Daten und zugehörigen Materialien eine der Kernfunktionen des Datenarchivs. Dies ist notwendig, da die durch digitale Objekte repräsentierten Informationen durch Einbußen in ihrer Integrität, Authentizität und Vertraulichkeit sowie den gänzlichen Verlust der Verfügbarkeit und Nutzbarkeit bedroht sind (Dobratz und Schoger 2010: 94). Auch wenn die grundsätzlichen Arbeitsabläufe lange vor der Etablierung des OAIS-Standards (ISO 14721:2003) entwickelt wurden, bildet dieser doch einen relevanten Bezugsrahmen und liefert eine Nomenklatur für die Darstellung. Wie eingangs erwähnt, lag und liegt der Fokus der Arbeit des Datenarchivs darauf, eine möglichst breite Nachnutzung der archivierten Forschungsdaten zu ermöglichen. Vor diesem Hintergrund ergab sich mehr oder weniger zwangsläufig auch die Notwendigkeit, Maßnahmen zum Erhalt des Datenbestandes durchzuführen. Maßnahmen, die heutzutage unter dem Begriff „Langzeitarchivierung“ diskutiert werden.

Nach Abschluss der oben unter Ingest und Aufbereitung/Dokumentation beschriebenen Prozesse erfolgt die Archivierung im engeren Sinne durch die Erzeugung des Archival Information Packages oder Archivinformationspaketes (AIP)²⁸ und dessen Überführung in den zentralen Archivspeicher. Die Zusammensetzung des AIP variiert dabei je nach Aufbereitungs-, Dokumentations- und Publikationsziel einer Studie. Es enthält unter anderem:

26 <http://www.gesis.org/unser-angebot/daten-analysieren/soziale-indikatoren/simon-social-indicators-monitor/>

27 <http://www.gesis.org/unser-angebot/daten-analysieren/umfragedaten/>

28 „Ein Informationspaket, bestehend aus der Inhaltsinformation und den dazugehörigen Erhaltungsmetadaten, das innerhalb eines OAIS aufbewahrt wird.“ (nestor 2012: 8)

- vom Datengeber übernommene Original-Objekte (zum Beispiel Datensätze, Dokumente) bzw. Kopien davon (SIP)
- im Archiv aufbereitete Datensatzversionen
- Aufbereitungssyntax, die den Bezug zwischen Original- und Archivversion herstellen, sowie weitere Dokumentation der Aufbereitung
- im Archiv erzeugte Metadaten auf Variablen- und Studienebene (im Archiv erstellte DDI-Instanzen) zur Beschreibung der Daten sowie weitere technische und administrative Metadaten.

Zusätzlich zu dem oben beschriebenen AIP werden die im Verlauf des Ingestprozesses ebenfalls erzeugten Dissemination Information Packages im Studienarchiv abgelegt, also die Objekte des AIP, die Nutzern über verschiedene Services in geeigneten Formaten zur Verfügung gestellt werden (OAIS: Access Funktion).

Organisation des Archivspeichers

Der Archivspeicher ist in Form einer dateibasierten Verzeichnisstruktur organisiert. Alle zu einer Studie gehörenden digitalen Objekte (AIP und DIP) werden dort so abgelegt, dass sie den Archiv-Lebenszyklus einer Studie reflektieren. Die Dateien werden nach bestimmten Regeln in definierte Verzeichnisse abgelegt und nach einem einheitlichen Schema benannt (die Originaldateien behalten ihren ursprünglichen Namen). Durch dieses Vorgehen werden die technischen und logischen Beziehungen der Objekte zueinander ausgedrückt.

Das Studienarchiv unterliegt restriktiven Zugriffsrechten. Ein erweiterter Kreis der Archivmitarbeiter hat zwar Leserechte, jedoch besitzen nur zwei Personen das Recht, Dateien hinzuzufügen, zu löschen oder zu ändern. An verschiedenen Punkten im Workflow werden digitale Objekte erzeugt, die der Langfristsicherung zugeführt werden müssen. Der Austausch erfolgt über ein spezielles Verzeichnis außerhalb des digitalen Archivs. Der für den Archivspeicher zuständige Mitarbeiter übernimmt von dort alle anfallenden Objekte, überprüft sie auf die Einhaltung der Langfristsicherungsstandards (Formate, Namenskonventionen etc.), passt sie ggf. an und überführt sie dann an die entsprechende Stelle im Archivverzeichnis. Durch diesen Ablauf wird sichergestellt, dass nur autorisierte Personen langfristsicherungsfähige Objekte in das digitale Archiv überführen bzw. aus dem Archiv entfernen können.

Physische Speicherung und Erhaltung der Archivdaten

Neben den zuvor beschriebenen organisatorischen und technischen Maßnahmen und dem Vorhalten einer modernen und leistungsfähigen IT-Infrastruktur setzt das GESIS Datenarchiv weitere Maßnahmen ein, die zur Sicherstellung des physischen Erhalts von archivierten Objekten geeignet sind und im Desasterfall eine schnelle Wiederherstellung ermöglichen. Dazu gehören die üblichen Maßnah-

men und Vorkehrungen, die Teil eines aktuellen IT-Sicherheitskonzeptes sein sollten. Dies reicht von grundlegenden Maßnahmen zum physischen Schutz vor Ort (wie dauerhaft verschlossene Rechenzentren, die nur für Berechtigte zugänglich sind, über Rauch- und Wassermeldeanlage, Temperaturüberwachung, unterbrechungsfreie Stromversorgung etc.), räumlich getrennter und redundanter Datenhaltung (onsite, off site, offline), Diversifizierung der eingesetzten Speichertechnik sowie regelmäßige Medienmigration.

Sicherstellung der langfristigen Nutzbarkeit und Interpretierbarkeit

Diese technischen Maßnahmen alleine reichen allerdings nicht aus, um die Nutzbarkeit der Datenbestände langfristig zu erhalten, da sie primär auf die physische Sicherung der digitalen Objekte in ihrem jeweiligen Zustand ausgerichtet sind (Bitstream-Preservation). Die weitaus größere Herausforderung für das Langzeitarchiv ist es aber, die Daten, Metadaten und Dokumente so zu erhalten, dass deren Lesbarkeit bzw. Interpretierbarkeit auch in der Zukunft garantiert werden kann. Insbesondere die Bedrohung der digitalen Bestände durch technischen Fortschritt bzw. dessen Kehrseite (Hardware und Speichermedien, die nicht mehr unterstützt oder produziert werden, sowie veraltete Betriebssysteme, Applikationen und Dateiformate) ist eine Konstante in der über fünfzigjährigen Archivarbeit. Der Erhalt der Interpretierbarkeit der archivierten Studien wird in der Hauptsache durch Migrationsstrategien erreicht. Bei Bedarf wurden in der Vergangenheit auch Emulationen bzw. Virtualisierungen als mittelfristige Lösung eingesetzt, bis die Bestände komplett migriert werden konnten.²⁹

Die langfristige Nutzbarkeit des Archivbestandes wird insbesondere durch folgende Maßnahmen gesichert:

1. Die technische Entwicklung der im Archiv eingesetzten Speichertechnik, Medien, Software und insbesondere der damit verbundenen Dateiformate wird kontinuierlich verfolgt.
2. Bei der Aufnahme von Studien ins Langzeitarchiv werden Daten und Dokumente in definierte und standardisierte Formate überführt. Diese Maßnahme führt zu einer Reduktion der im Archiv vorhandenen Formate, was einerseits das Verfolgen der technischen Entwicklungen

²⁹ Als zentrale Produktions- und Archivierungsplattform (Aufbereitung, Dokumentation, Archivierung) des Datenarchivs diente bis Ende der 90er Jahre eine IBM Mainframe. Nach und nach wurde die Datenaufbereitung und -dokumentation sowie ein Großteil des Archivbestandes in zeitgemäßere Arbeitsumgebungen migriert. Da der Betrieb des Großrechners hohe Wartungs- und Unterhaltungskosten verursachte, wurde die Hardware abgeschafft und der Betrieb in eine virtuelle Umgebung verlagert. Die Migration der noch auf Mainframe-Medien (Magnetspulen und -kassetten, Festplatten) befindlichen Archivbestände konnte dann mit einem deutlich niedrigeren Zeitdruck realisiert werden.

erleichtert und andererseits im Fall einer Migration, den dafür notwendigen Ressourcenbedarf deutlich verringert.

3. Wenn Bestände von technischen Entwicklungen bedroht sind oder sie in der vorliegenden Form nicht mehr zeitgemäß verarbeitet werden können, werden Strategien zur Migration entwickelt und umgesetzt. Je nach Notwendigkeit setzen dabei die konkreten Maßnahmen auf unterschiedlichen Ebenen an: Reine Datenträgermigrationen (sprich, die Kopie von einem Medium auf ein anderes) finden im Prinzip kontinuierlich statt. Formatmigrationen, da ungleich aufwendiger, werden nur dann vorgenommen, wenn die Gefahr besteht, dass die archivierten Objekte aufgrund des Formates in Zukunft nicht mehr nutzbar sind oder wenn mit der Migration so große Vorteile für die Nutzung oder die Archivarbeit einhergehen, dass der Aufwand zu rechtfertigen ist.³⁰ Dabei wird sichergestellt, dass die wesentlichen Inhalte der Objekte nicht verändert werden, und durch entsprechende Verfahren und Dokumentation der Prozesse, die Authentizität der migrierten Objekte erhalten bzw. wiederhergestellt.

Ausblick

In der letzten Dekade hat sich die (sozialwissenschaftliche) Datenlandschaft sehr dynamisch entwickelt. Eine Vielzahl neuer Datenserviceeinrichtungen sind entstanden, wie etwa die im Umfeld des RatSWD angesiedelten Forschungsdatenzentren³¹, wichtige, auf Dauer angelegte Erhebungsprogramme wurden etabliert, die rund um diese komplexen Studien eigene Daten(zugangs)infrastrukturen aufgebaut haben (zum Beispiel NEPS³², SHARE³³) und immer mehr – auch kleinere – Forschungsprojekte stellen ihre Forschungsdaten selbstständig für eine Nachnutzung zur Verfügung. Darüber hinaus verändern sich auch die Daten selbst: Komplexere Forschungsdesigns, neue Formen von Daten und insbesondere auch neue (technische) Möglichkeiten der Verbindung verschiedener Datenquellen und -formen – auch über Disziplingrenzen hinweg – werden zunehmend wichtiger. Insgesamt hat das Thema Forschungsdaten enorm an Bedeutung

³⁰ Eines der größten Migrationsprojekte der letzten Jahre betraf den umfangreichen Bestand an sogenannten Codebüchern. Neben der Dokumentation auf Studienebene beinhalten diese eine vollständige Dokumentation der Variablen des Datensatzes (inklusive der vollständigen Frage- und Antworttexte). In einem mehrjährigen Projekt wurde dabei die in den Codebüchern enthaltene Dokumentation in den XML-basierten Metadatenstandard der Data Documentation Initiative (DDI, www.ddialliance.org) und somit in ein auch für die Langfristsicherung geeignetes Format überführt.

³¹ <http://www.ratswd.de/dat/fdz.php>

³² <http://www.neps-data.de>

³³ <http://www.share-project.org>

gewonnen und damit verbundene Fragen beispielsweise der Zugänglichkeit, des Datenschutzes und auch der Langzeitarchivierung rücken in den Fokus der Aufmerksamkeit. Gleichzeitig wachsen die Anforderungen an Forschende bzw. Forschungsprojekte bezüglich des Umgangs mit den von ihnen erzeugten Daten. Das Erstellen von Datenmanagementplänen bzw. ganz allgemein, die Adressierung von Fragen des zukünftigen Umgangs mit generierten Forschungsdaten wird mehr und mehr Bestandteil der regulären Projektplanung und spielt zunehmend auch bei der Beantragung von Forschungsgeldern eine wichtige Rolle.³⁴

Diese Entwicklungen greift das Datenarchiv bereits seit einiger Zeit in unterschiedlichen Projekten und Initiativen auf. Insbesondere mit Blick auf die zunehmend verteilte Dateninfrastruktur und die damit auch tendenziell einhergehende Unübersichtlichkeit aus Nutzersicht hat GESIS seine Aktivitäten im Bereich strukturierender und übergreifender Dienste deutlich verstärkt. So schafft die von GESIS gegründete und mittlerweile gemeinsam mit der Zentralbibliothek Wirtschaftswissenschaften (ZBW) betriebene Registrierungsagentur für Sozial- und Wirtschaftsdaten (da|ra) die Voraussetzungen für eine dauerhafte Identifizierung, Sicherung und Lokalisierung von Forschungsdaten. Damit wird nicht nur die Grundlage für eine zuverlässige und dauerhafte Zitierbarkeit von Daten geschaffen und Datenproduzenten auch ein Reputationsgewinn aus solchen Zitationen ermöglicht, sondern es gehen damit auch deutlich verbesserte Möglichkeiten einher, Publikationen und zugrundeliegende Daten effizienter zu verbinden. Desweiteren arbeitet GESIS im Kontext von da|ra seit Ende 2011 am Aufbau eines umfassenden Datennachweissystems, das existierende Datenbestände auch über die bei da|ra registrierten hinaus beschreibt, auf die entsprechenden Datenquellen verweist und Hinweise über die Zugänglichkeit gibt. Dieser zentrale Nachweis sozialwissenschaftlicher Datenbestände, der in einer späteren Ausbaustufe auch noch um eine Selbstmeldekomponente erweitert wird, soll dazu beitragen, die an vielen verschiedenen Stellen erzeugten und zunehmend auch zur Nachnutzung bereitstehenden Datenbestände für die Wissenschaft sichtbar und damit auch nutzbar zu machen.

Mit dem kürzlich gegründeten 'Archiving and Data Management Training and Information Center'³⁵ fördert GESIS systematisch das Bewusstsein für die Wichtigkeit guten Datenmanagements und der Bedeutung einer professionellen Archivierung und Pflege (data curation) über den gesamten Lebenszyklus von Daten. Entsprechende Kenntnisse werden durch Trainingskurse, gezielte Beratung und die Bereitstellung von Informationsangeboten vermittelt. Die Angebote des Training Centers richten sich dabei sowohl an Wissenschaftler und Projekte,

34 Für eine ausführliche Darstellung zur Bedeutung von Datenmanagementplänen und ihrem Aufbau siehe Jensen (2011).

35 <http://www.gesis.org/en/archive-and-data-management-training-and-information-centre/training-center-home/>

die Daten generieren bzw. planen dies künftig zu tun, als auch an Experten aus Datenserviceeinrichtungen, Archiven oder Bibliotheken, um diese im Hinblick auf ihre Tätigkeit weiter zu professionalisieren.

Angesichts der eingangs dargestellten Entwicklungen bedarf aber auch die eigentliche Archivierungstätigkeit des GESIS Datenarchivs im engeren Sinne einer stetigen Anpassung und Weiterentwicklung. Das Datenarchiv bietet mittels eines relativ komplexen Workflows ein weites Spektrum häufig sehr eng miteinander verbundener Dienstleistungen an. Um aber auch solchen Projekten oder Institutionen bedarfsgerechte Angebote machen zu können, die nur einzelne oder mehrere Komponenten aus dem Dienstleistungsportfolio in Anspruch nehmen wollen, beispielsweise Langzeitarchivierungsdienste, sollen zukünftig verstärkt separate, aber standardisierte Dienstleistungen angeboten werden, die dann (weitgehend frei) kombinierbar sein sollen. In diesem Zusammenhang sollen insbesondere Hintergrunddienste für andere Datenzentren im Bereich der Langzeitarchivierung angeboten werden. Darüber hinaus wird das Dienstleistungsangebot des Archivs Anfang 2013 um einen neuen Service erweitert, nämlich die vom Nutzer (weitgehend) eigenständig durchgeführte Publikation und Distribution von Daten mittels eines Daten-Repositoriums (DATORIUM). Mit diesem Angebot reagiert GESIS auf den Bedarf von Datenproduzenten und Forschern nach schnellen und flexiblen Publikations- und Distributionswegen, die es ihnen insbesondere erlauben, Forschungsergebnisse mit anderen Wissenschaftler/inne/n zu teilen und diesen somit auch Sichtbarkeit zu geben. Auch im Hinblick auf den Zugang zu datenschutzrechtlich sensibleren Varianten der bei GESIS gehaltenen Daten wird es in 2013 eine Erweiterung des Services geben. Bisher stellt das Datenarchiv Daten in der Regel als vollständig oder faktisch anonymisierte Public bzw. Scientific Use Files zur Verfügung. Die dafür notwendigen Anonymisierungsmaßnahmen sind in Einzelfällen nahezu unmöglich (zum Beispiel bei Elitestudien) oder können auch teilweise mit einer erheblichen Einschränkung des Analysepotentials für bestimmte Auswertungszwecke einhergehen (zum Beispiel durch die Löschung ganzer Variablen oder identifizierbarer Fälle oder durch Vergrößerung oder Zufallsüberlagerung bestimmter zur Re-Identifikation geeigneter Merkmale wie Berufsklassen oder kleinräumige regionale Identifikatoren). Um zukünftig auch die Zugänglichkeit solcher Daten für wissenschaftliche Forschungszwecke systematisch zu verbessern, baut GESIS derzeit ein Secure Data Center auf, das in einer ersten Ausbaustufe für schutzwürdige Datenbestände Vor-Ort-Zugänge an den GESIS Standorten ermöglicht. In einer weiteren Ausbaustufe soll dieser Zugang dann auch um die Möglichkeit eines Fernzugriffes erweitert werden.

Eine der größten Herausforderungen für die Zukunft geht für das Datenarchiv jedoch sicher mit dem Aufkommen neuer Datenformen und -typen einher, deren Archivierung teilweise völlig neue Anforderungen an das Datenarchiv stellen würde.

Literatur

- Dobratz, S. und Schoger, A. (2010): Grundkonzepte der Vertrauenswürdigkeit. In: Neuroth, H./Oßwald, A./Scheffel, R./Strathmann, S. und Huth, K. (Hrsg.): nestor-Handbuch: Eine kleine Enzyklopädie der digitalen Langzeitarchivierung. Version 2.3 – 2010, Kapitel 5:2 (ohne Seitenzahl).
- Jensen, U. (2011): Datenmanagementpläne. In: Büttner, S./Hobohm, H.-C. und Müller, L. (Hrsg.): Handbuch Forschungsdatenmanagement. Bad Honnef: Bock + Herchen, 71-82.
- JISC (Joint Information Systems Committee) (2006): Digital Preservation. Continued access to authentic digital assets. Briefing paper 20. November 2006, 1. <http://www.jisc.ac.uk/media/documents/publications/digitalpreservationbp.pdf> [21.05.2012]
- Kleiner, M. (2012): Vorwort. In: Neuroth, H./Strathmann S./Oßwald, A./Scheffel, R./Klump, J. und Ludwig, J. (Hrsg.): Langzeitarchivierung von Forschungsdaten. Eine Bestandsaufnahme. Boizenburg: vwh, 9-13.
- nestor-Arbeitsgruppe OAI-S-Übersetzung / Terminologie (Hrsg.) (2012): Referenzmodell für ein Offenes Archiv-Informations-System – Deutsche Übersetzung. nestor-Materialien 16. Frankfurt am Main: nestor. urn:nbn:de:0008-2012051101.
- nestor-Arbeitsgruppe Vertrauenswürdige Archive – Zertifizierung (Hrsg.) (2006): Kriterienkatalog vertrauenswürdige digitale Langzeitarchive. Version 1 (Entwurf zur Öffentlichen Kommentierung). nestor-Materialien 8. Frankfurt am Main: nestor. urn:nbn:de:0008-2006060710.
- Quandt, M. und Mauer, R. (2012): Sozialwissenschaften. In: Neuroth, H./Strathmann, S./Oßwald, A./Scheffel, R./Klump, J. und Ludwig, J. (Hrsg.): Langzeitarchivierung von Forschungsdaten. Eine Bestandsaufnahme. Boizenburg: vwh, 61-81.
- Scheuch, E.K. (1990): From a data archive to an infrastructure for the social sciences. International Social Science Journal (123), 93-111.
- Scheuch, E.K. (2003): History and visions in the development of data services for the social sciences. International Social Science Journal (177), 385-399.
- Wagner, G. und Huschka, D. (2012): Datenverfügbarkeit reicht nicht, um Replikationsstudien zur Routine zu machen. Working Paper No. 194. RatSWD Working Paper Series Januar 2012. Berlin: RatSWD.
- Zenk-Möltgen, W. und Habbel, N. (2012): Der GESIS Datenbestandskatalog und sein Metadatenschema. Version 1.8. GESIS Technical Reports 2012-01.